

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
24 June 2004 (24.06.2004)

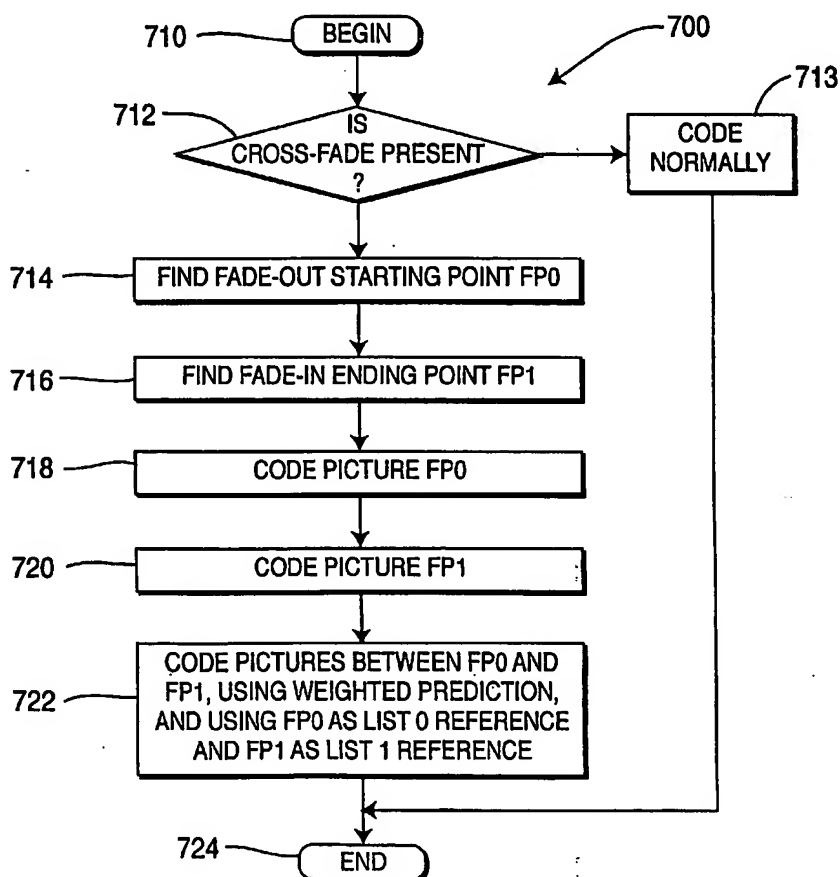
PCT

(10) International Publication Number  
**WO 2004/054225 A2**

- (51) International Patent Classification<sup>7</sup>: H04N (72) Inventor; and  
(21) International Application Number: PCT/US2003/036413 (75) Inventor/Applicant (for US only): BOYCE, Jill, Mac-Donald [US/US]; 3 Brandywine Court, Manalapan, NJ 07726 (US).  
(22) International Filing Date: 13 November 2003 (13.11.2003) (74) Agents: TRIPOLI, Joseph, S. et al.; c/o Thomson Licensing Inc., Two Independence Way, Suite #200, Princeton, NJ 08540 (US).  
(25) Filing Language: English (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.  
(26) Publication Language: English (84) Designated States (regional): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
(30) Priority Data: 60/430,793 4 December 2002 (04.12.2002) US  
(71) Applicant (for all designated States except US): THOMSON LICENSING S.A. [FR/FR]; 46, Quai A. Le Gallo, F-92648 Boulogne (FR).

[Continued on next page]

(54) Title: ENCODING OF VIDEO CROSS-FADES USING WEIGHTED PREDICTION



(57) Abstract: A video encoder (200, 300) and method (700) are provided for encoding video signal data for at least one cross-fade picture disposed between a fade-out start picture and a fade-in end picture, where the encoder portion includes a reference picture weighting factor unit (272, 372) for assigning weighting factors corresponding to each of the fade-out start picture and the fade-in end picture, respectively, and the method for encoding cross-fades between pictures includes identifying pictures between which a cross-fade is desired, determining (714,716) appropriate end-points for the cross-fade, and encoding (718,720) the end-points prior to encoding (722) the cross-fade picture.



Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,  
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE,  
SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA,  
GN, GQ, GW, ML, MR, NE, SN, TD, TG).

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**Published:**

- *without international search report and to be republished upon receipt of that report*

## ENCODING OF VIDEO CROSS-FADES USING WEIGHTED PREDICTION

### CROSS-REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. Provisional Application Serial No. 60/430,793 (Attorney Docket No. PU020487), filed December 4, 2002 and entitled "ENCODING OF VIDEO CROSS-FADES USING WEIGHTED PREDICTION", which is incorporated herein by reference in its entirety.

### FIELD OF THE INVENTION

The present invention is directed towards video encoders, and in particular, towards an apparatus and method for effectively producing video cross-fades between pictures.

### BACKGROUND OF THE INVENTION

Video data is generally processed and transferred in the form of bit streams. Typical video compression coders and decoders ("CODECs") gain much of their compression efficiency by forming a reference picture prediction of a picture to be encoded, and encoding the difference between the current picture and the prediction. The more closely that the prediction is correlated with the current picture, the fewer bits that are needed to compress that picture, thereby increasing the efficiency of the process. Thus, it is desirable for the best possible reference picture prediction to be formed.

In many video compression standards, including Moving Picture Experts Group ("MPEG")-1, MPEG-2 and MPEG-4, a motion compensated version of a previous reference picture is used as a prediction for the current picture, and only the difference between the current picture and the prediction is coded. When a single picture prediction ("P" picture) is used, the reference picture is not scaled when the motion compensated prediction is formed. When bi-directional picture predictions ("B" pictures) are used, intermediate predictions are formed from two different pictures, and then the two intermediate predictions are averaged together, using equal weighting factors of  $(\frac{1}{2}, \frac{1}{2})$  for each, to form a single averaged prediction.

In some video sequences, in particular those with fades, the current picture to be coded is more strongly correlated to the reference picture scaled by a weighting

factor than to the reference picture itself. The Joint Video Team ("JVT") video compression standard allows weighting factors and offsets to be sent for each reference picture. The standard specifies how the decoder will use the weighting factors, but it does not specify how an encoder might determine an appropriate weighting factor. For sequences that include cross-fades, determining the appropriate weighting factors and reference pictures to use is quite difficult.

### SUMMARY OF THE INVENTION

These and other drawbacks and disadvantages of the prior art are addressed by an apparatus and method that efficiently compress video cross-fades using JVT weighted prediction. The end-points of a cross-fade are determined and used as reference pictures for encoding pictures in the cross-fade region.

An apparatus and method are provided for encoding video signal data for a cross-fade picture disposed between a fade-out or start picture and a fade-in or end picture, where the encoder portion includes a reference picture weighting factor unit for assigning weighting factors corresponding to each of the fade-out start picture and the fade-in end picture, respectively, and the method for encoding cross-fades between pictures includes identifying pictures between which a cross-fade is desired, determining appropriate end-points for the cross-fade, and encoding the end-points prior to encoding the cross-fade picture.

These and other aspects, features and advantages of the present disclosure will become apparent from the following description of exemplary embodiments, which is to be read in connection with the accompanying drawings.

### BRIEF DESCRIPTION OF THE DRAWINGS

The present invention may be better understood with reference to the following exemplary figures, in which:

Figure 1 shows a block diagram for a standard video encoder;

Figure 2 shows a block diagram for a video encoder with implicit reference picture weighting for video cross-fades;

Figure 3 shows a block diagram for a video encoder with explicit reference picture weighting for video cross-fades;

Figure 4 shows a block diagram for a video decoder with explicit reference picture weighting for video cross-fades;

Figure 5 shows a pictorial representation of a video cross-fade between a pair of pictures; and

5        Figure 6 shows a flowchart for an exemplary encoding process.

#### DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

An apparatus and method are disclosed for encoding of video cross-fades using weighted prediction, including motion vector estimation and adaptive reference  
10        picture weighting factor assignment. In some video sequences, in particular those with fading, the current picture or image block to be coded is more strongly correlated to a reference picture scaled by a weighting factor than to the reference picture itself. Video encoders without weighting factors applied to reference pictures encode fading sequences very inefficiently. When weighting factors are used in encoding, a video  
15        encoder needs to determine both weighting factors and motion vectors, but the best choice for each of these depends on the other.

Hence, a method is described to efficiently compress video cross-fades using JVT weighted prediction. The end-points of a cross-fade are first determined and used as the reference pictures for encoding the pictures in the cross-fade region.

20        The present description illustrates the principles of the invention. It will thus be appreciated that those skilled in the art will be able to devise various arrangements that, although not explicitly described or shown herein, embody the principles of the invention and are included within its spirit and scope.

All examples and conditional language recited herein are intended for  
25        pedagogical purposes to aid the reader in understanding the principles of the invention and the concepts contributed by the inventor to furthering the art, and are to be construed as being without limitation to such specifically recited examples and conditions.

Moreover, all statements herein reciting principles, aspects, and embodiments  
30        of the invention, as well as specific examples thereof, are intended to encompass both structural and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents as well as equivalents

developed in the future, i.e., any elements developed that perform the same function, regardless of structure.

Thus, for example, it will be appreciated by those skilled in the art that the block diagrams herein represent conceptual views of illustrative circuitry embodying the principles of the invention. Similarly, it will be appreciated that any flow charts, flow diagrams, state transition diagrams, pseudocode, and the like represent various processes which may be substantially represented in computer readable media and so executed by a computer or processor, whether or not such computer or processor is explicitly shown.

The functions of the various elements shown in the figures may be provided through the use of dedicated hardware as well as hardware capable of executing software in association with appropriate software. When provided by a processor, the functions may be provided by a single dedicated processor, by a single shared processor, or by a plurality of individual processors, some of which may be shared.

Moreover, explicit use of the term "processor" or "controller" should not be construed to refer exclusively to hardware capable of executing software, and may implicitly include, without limitation, digital signal processor ("DSP") hardware, read-only memory ("ROM") for storing software, random access memory ("RAM"), and non-volatile storage.

Other hardware, conventional and/or custom, may also be included. Similarly, any switches shown in the figures are conceptual only. Their function may be carried out through the operation of program logic, through dedicated logic, through the interaction of program control and dedicated logic, or even manually, the particular technique being selectable by the implementer as more specifically understood from

the context.

In the claims hereof, any element expressed as a means for performing a specified function is intended to encompass any way of performing that function including, for example, a) a combination of circuit elements that performs that function or b) software in any form, including, therefore, firmware, microcode or the like, combined with appropriate circuitry for executing that software to perform the function. The invention as defined by such claims resides in the fact that the functionalities provided by the various recited means are combined and brought

together in the manner which the claims call for. Applicant thus regards any means that can provide those functionalities as equivalent to those shown herein.

In some video sequences, in particular those with fading, the current picture or image block to be coded is more strongly correlated to a reference picture scaled by a weighting factor than to the reference picture itself. Video encoders without weighting factors applied to reference pictures encode fading sequences very inefficiently.

In the Joint Video Team ("JVT") video compression standard, each P picture can use multiple reference pictures to form a picture's prediction, but each individual macroblock or macroblock partition (of size 16x8, 8x16, or 8x8) uses only a single reference picture for prediction. In addition to coding and transmitting the motion vectors, a reference picture index is transmitted for each macroblock or macroblock partition, indicating which reference picture is used. A limited set of possible reference pictures is stored at both the encoder and decoder, and the number of allowable reference pictures is transmitted. Unlike in previous standards, such as MPEG-2, a JVT encoder has considerable flexibility in that previously coded pictures can be used as reference pictures.

In the JVT standard for bi-predictive pictures (also called "B" pictures), two predictors are formed for each macroblock or macroblock partition, each of which can be from a separate reference picture, and the two predictors are averaged together to form a single averaged predictor. For bi-predictively coded motion blocks, the reference pictures can both be from the forward direction, both be from the backward direction, or one each from the forward and backward directions.

Two lists are maintained of the available reference pictures that may be used for prediction. The two reference pictures are referred to as the List 0 and List 1 predictors. An index for each reference picture is coded and transmitted, `ref_idx_l0` and `ref_idx_l1`, for the List 0 and List 1 reference pictures, respectively.

The JVT standard provides two modes of weighted prediction, which allows weighting factors and/or offsets to be applied to reference pictures when forming a prediction. The weighting factor to be used is based on the reference picture index (or indices in the case of bi-prediction) for the current macroblock or macroblock partition. The reference picture indices are either coded in the bitstream or may be derived, such as for skipped or direct mode macroblocks. A single weighting factor

and single offset are associated with each reference picture index for all of the slices of the current picture. In explicit mode, these parameters are coded in the slice header. In implicit mode, these parameters are derived. The weighting factors and offset parameter values are constrained to allow for 16-bit arithmetic operations in the inter-prediction process. The encoder may select either implicit mode or explicit mode for each coded picture.

JVT bi-predictive or "B" pictures allow adaptive weighting between the two predictions, i.e.,  $Pred = [(P0) * (Pred0)] + [(P1) * (Pred1)] + D$ , where P0 and P1 are weighting factors, Pred0 and Pred1 are the reference picture predictions for List 0 and List 1 respectively, and D is an offset.

As shown in Figure 1, a standard video encoder is indicated generally by the reference numeral 100. An input to the encoder 100 is connected in signal communication with a non-inverting input of a summing junction 110. The output of the summing junction 110 is connected in signal communication with a block transform function 120. The transform 120 is connected in signal communication with a quantizer 130. The output of the quantizer 130 is connected in signal communication with a variable length coder ("VLC") 140, where the output of the VLC 140 is an externally available output of the encoder 100.

The output of the quantizer 130 is further connected in signal communication with an inverse quantizer 150. The inverse quantizer 150 is connected in signal communication with an inverse block transformer 160, which, in turn, is connected in signal communication with a reference picture store 170. A first output of the reference picture store 170 is connected in signal communication with a first input of a motion estimator 180. The input to the encoder 100 is further connected in signal communication with a second input of the motion estimator 180. The output of the motion estimator 180 is connected in signal communication with a first input of a motion compensator 190. A second output of the reference picture store 170 is connected in signal communication with a second input of the motion compensator 190. The output of the motion compensator 190 is connected in signal communication with an inverting input of the summing junction 110.

Turning to Figure 2, a video encoder with implicit reference picture weighting is indicated generally by the reference numeral 200. An input to the encoder 200 is connected in signal communication with a non-inverting input of a summing junction



210. The output of the summing junction 210 is connected in signal communication with a block transformer 220. The transformer 220 is connected in signal communication with a quantizer 230. The output of the quantizer 230 is connected in signal communication with a VLC 240, where the output of the VLC 240 is an  
5 externally available output of the encoder 200.

The output of the quantizer 230 is further connected in signal communication with an inverse quantizer 250. The inverse quantizer 250 is connected in signal communication with an inverse block transformer 260, which, in turn, is connected in signal communication with a reference picture store 270. A first output of the  
10 reference picture store 270 is connected in signal communication with a first input of a reference picture weighting factor assignor 272. The input to the encoder 200 is further connected in signal communication with a second input of the reference picture weighting factor assignor 272. A second output of the reference picture store 270 is connected in signal communication with a second input of the motion estimator  
15 280.

The input to the encoder 200 is further connected in signal communication with a third input of the motion estimator 280. The output of the motion estimator 280, which is indicative of motion vectors, is connected in signal communication with a first input of a motion compensator 290. A third output of the reference picture store 270  
20 is connected in signal communication with a second input of the motion compensator 290. The output of the motion compensator 290, which is indicative of a motion compensated reference picture, is connected in signal communication with a first input of a multiplier or reference picture weighting applicator 292. Although an exemplary multiplier embodiment is shown, the reference picture weighting applicator  
25 292 may be implemented in alternate ways, such as, for example, by a shift register. The output of the reference picture weighting factor assignor 272, which is indicative of a weighting factor, is connected in signal communication with a second input of the reference picture weighting applicator 292. The output of the reference picture weighting applicator 292 is connected in signal communication with an inverting input  
30 of the summing junction 210.

Turning to Figure 3, a video encoder with explicit reference picture weighting is indicated generally by the reference numeral 300. An input to the encoder 300 is connected in signal communication with a non-inverting input of a summing junction

310. The output of the summing junction 310 is connected in signal communication with a block transformer 320. The transformer 320 is connected in signal communication with a quantizer 330. The output of the quantizer 330 is connected in signal communication with a VLC 340, where the output of the VLC 340 is an  
5 externally available output of the encoder 300.

The output of the quantizer 330 is further connected in signal communication with an inverse quantizer 350. The inverse quantizer 350 is connected in signal communication with an inverse block transformer 360, which, in turn, is connected in signal communication with a reference picture store 370. A first output of the  
10 reference picture store 370 is connected in signal communication with a first input of a reference picture weighting factor assignor 372. The input to the encoder 300 is further connected in signal communication with a second input of the reference picture weighting factor assignor 372. A first output of the reference picture weighting factor assignor 372, which is indicative of a weighting factor, is connected in signal  
15 communication with a first input of a motion estimator 380. A second output of the reference picture store 370 is connected in signal communication with a second input of the motion estimator 380.

The input to the encoder 300 is further connected in signal communication with a third input of the motion estimator 380. The output of the motion estimator 380,  
20 which is indicative of motion vectors, is connected in signal communication with a first input of a motion compensator 390. A third output of the reference picture store 370 is connected in signal communication with a second input of the motion compensator 390. The output of the motion compensator 390, which is indicative of a motion compensated reference picture, is connected in signal communication with a first  
25 input of a multiplier or reference picture weighting applicator 392. A second output of the reference picture weighting factor assignor 372, which is indicative of a weighting factor, is connected in signal communication with a second input of the reference picture weighting applicator 392. The output of the reference picture weighting applicator 392 is connected in signal communication with a first non-inverting input of  
30 a summing junction 394. A third output of the reference picture weighting factor assignor 372, which is indicative of an offset, is connected in signal communication with a second non-inverting input of the summing junction 394. The output of the

summing junction 394 is connected in signal communication with an inverting input of the summing junction 310.

As shown in Figure 4, a video decoder for explicit reference picture weighting is indicated generally by the reference numeral 500. The video decoder 500 includes  
5 a variable length decoder ("VLD") 510 connected in signal communication with an inverse quantizer 520. The inverse quantizer 520 is connected in signal communication with an inverse transformer 530. The inverse transformer 530 is connected in signal communication with a first input terminal of a summing junction 540, where the output of the summing junction 540 provides the output of the video  
10 decoder 500. The output of the summing junction 540 is connected in signal communication with a reference picture store 550. The reference picture store 550 is connected in signal communication with a motion compensator 560, which is connected in signal communication with a first input of a multiplier or reference picture weighting applicator 570. As will be recognized by those of ordinary skill in  
15 the pertinent art, the decoder 500 for explicit weighted prediction may also be used for implicit weighted prediction.

The VLD 510 is further connected in signal communication with a reference picture weighting factor lookup 580 for providing a coefficient index to the lookup 580. A first output of the lookup 580 is for providing a weighting factor, and is connected in  
20 signal communication to a second input of the reference picture weighting applicator 570. The output of the reference picture weighting applicator 570 is connected in signal communication to a first input of a summing junction 590. A second output of the lookup 580 is for providing an offset, and is connected in signal communication to a second input of the summing junction 590. The output of the summing junction 590  
25 is connected in signal communication with a second input terminal of the summing junction 540.

As shown in Figure 5, a picture cross-fade is indicated generally by the reference numeral 600. The exemplary picture cross-fade 600 includes a fade-out or starting picture 610, identified as FP0, and a fade-in or ending picture 612, identified  
30 as FP1.

Turning now to Figure 6, an exemplary process for encoding video signal data for an image block is indicated generally by the reference numeral 700. The process 700 is implemented with an encoder, such as the encoder 200 or 300 of Figures 2

and 3, respectively. The process 700 includes a start block 710 that passes control to a decision block 712. The decision block 712 determines whether a cross-fade is present, and, if none is present, passes control to a function block 713. The function block 713 performs normal encoding and passes control to an end block 724.

5        However, if the decision block 712 finds a cross-fade, it passes control to a function block 714. The function block 714 finds the fade-out starting point, FP0, and passes control to a function block 716, which finds the fade-in ending point FP1. The block 716 passes control to a function block 718, which codes the fade-out start picture FP0 and passes control to a function block 720. The block 720 codes the  
10        fade-in end picture FP1 and passes control to a function block 722.

      The function block 722, in turn, codes pictures disposed in display order between FP0 and FP1, using weighted prediction with the picture FP0 as the list 0 reference and the picture FP1 as the list 1 reference. The function block 722 passes control to the end block 724.

15        An authoring tool used for video cross-fades between a pair of pictures includes a video encoder, such as the encoder 200 of Figure 2, and operates on pre-stored video content. In addition to the uncompressed video content, some additional information may be available such as decision lists and editing splice points. The video encoder in an authoring tool does not necessarily need to operate in real time.  
20        Special effects such as fades and cross-fades can be applied in the authoring tool.

      Various techniques are well known for detecting fades and cross-fades, also known as dissolves, in video sequences. When encoding a particular picture, for each macroblock or macroblock partition, a JVT encoder must select a coding decision mode, one or two reference pictures, and one or more motion vectors.

25        ~~When a JVT encoder uses weighted prediction, once per picture or slice it may also select a weighting factor to be applied for each reference index used. One or more reference indices refer to each allowable reference picture, so multiple weights can be used for each individual reference picture.~~

      The authoring tool detects when a cross-fade is taking place. The authoring  
30        tool has sufficient information to detect when a cross-fade is taking place either because it applied the cross-fade itself, or because it read it from a decision list, or because it employs a fade detection algorithm. For a cross-fade, a picture identified as the fade-out starting point is identified as FP0 and the fade-in ending point picture

is identified as FP1. When a cross-fade is detected, the encoder codes pictures FP0 and FP1 prior to coding the pictures between FP0 and FP1 in display order, which are referred to as the cross-fade pictures. Thus, a feature of the present invention is that the fade-in end picture, FP1, is coded before the intermediate pictures.

5 It is common in video encoders to use a fixed pattern of the I, P and B picture coding types, and for the coding order to differ from the display order. For example, such a common pattern might comprise:

Common Coding Order: I0 P3 B1 B2 P6 B4 B5 P9 B7 B8

10 Common Display Order: I0 B1 B2 P3 B4 B5 P6 B7 B8 P9

For this common pattern, picture P3 is coded before the intermediate B1 and B2 pictures. The B1 and B2 pictures use I0 and P3 as reference picture in its prediction process.

15 The JVT standard does not require the use of fixed picture coding type patterns, and does not suggest methods by which an encoder can adjust the patterns to maximize coding efficiency. In accordance with the current invention, coding efficiency of cross-fading sequences can be improved by adjusting picture coding type and coding order. If, for example, picture 0 and picture 9 were identified as the  
20 fade-in start and fade-out end pictures, respectively, the following coding and display order could be used:

Inventive Coding Order: I0 P9 B1 B2 B3 B4 B5 B6 B7 B8

Inventive Display Order: I0 B1 B2 B3 B4 B5 B6 B7 B8 P9

25

When a cross-fade picture is encoded, the encoder orders the reference picture lists, using reference picture selection reordering if necessary, such that FP0 is the first picture on List 0 and FP1 is the first picture on List 1. This provides additional coding efficiency, because the reference index of 0, which refers to the first  
30 picture in the reference picture list, can be coded using a lesser number of bits than other reference indices. Then a weighting-factor is selected for the reference indices corresponding to each of FP0 and FP1, based on the relative contribution of the first picture and the second picture in the composition of the current picture. If the formula

used in creating the cross-fade picture is known, either because the authoring tool created the cross-fade, or from side-information, then the weighting factor from the composition formula can be used. If the exact formula is not known, a weighting factor can be computed using any of several different algorithms, such as those  
5 based on relative distance of the current picture from FP0 and FP1, for example.

This above described algorithm can be applied for all coded pictures in the cross-fade region, or may be applied only for those pictures that are marked to be stored as reference pictures. In alternate embodiments, either implicit mode or explicit mode weighted prediction may be used to code the cross-fade pictures. When  
10 explicit mode is used, any weighting factors may be used. When implicit mode is used, the weighting factors depend on the relative distance of the current picture from FP0 and FP1.

This system and technique may be applied to either Predictive "P" pictures, which are encoded with a single predictor, or to Bi-predictive "B" pictures, which are  
15 encoded with two predictors. The decoding processes, which are present in both encoder and decoders, are described below for the P and B picture cases. Alternatively, this technique may also be applied to coding systems using the concepts similar to I, B, and P pictures.

The same weighting factors can be used for single directional prediction in B  
20 pictures and for bi-directional prediction in B pictures. When a single predictor is used for a macroblock, in P pictures or for single directional prediction in B pictures, a single reference picture index is transmitted for the block. After the decoding process step of motion compensation produces a predictor, the weighting factor is applied to predictor. The weighted predictor is then added to the coded residual, and clipping is  
25 performed on the sum, to form the decoded picture. For use for blocks in P pictures or for blocks in B pictures that use only List 0 prediction, the weighted predictor is formed as:

$$\text{Pred} = W0 * \text{Pred0} + D0 \quad (1)$$

30

where W0 is the weighting factor associated with the List 0 reference picture, D0 is the offset associated with the List 0 reference picture, and Pred0 is the motion-compensated prediction block from the List 0 reference picture.

For use for blocks in B pictures that use only List 1 prediction, the weighted predictor is formed as:

$$\text{Pred} = W1 * \text{Pred1} + D1 \quad (2)$$

5

where W1 is the weighting factor associated with the List 1 reference picture, D1 is the offset associated with the List 1 reference picture, and Pred1 is the motion-compensated prediction block from the List 1 reference picture.

The weighted predictors may be clipped to guarantee that the resulting values will be within the allowable range of pixel values, typically 0 to 255. The precision of the multiplication in the weighting formulas may be limited to any pre-determined number of bits of resolution.

In the bi-predictive case, reference picture indexes are transmitted for each of the two predictors. Motion compensation is performed to form the two predictors. Each predictor uses the weighting factor associated with its reference picture index to form two weighted predictors. The two weighted predictors are then averaged together to form an averaged predictor, which is then added to the coded residual..

For use for blocks in B pictures that use List 0 and List 1 predictions, the weighted predictor is formed as:

20

$$\text{Pred} = (P0 * \text{Pred0} + D0 + P1 * \text{Pred1} + D1) / 2 \quad (3)$$

Clipping may be applied to the weighted predictor or any of the intermediate values in the calculation of the weighted predictor to guarantee that the resulting values will be within the allowable range of pixel values, typically 0 to 255.

Thus, a weighting factor is applied to the reference picture prediction of a video compression encoder and decoder that uses multiple reference pictures. The weighting factor adapts for individual motion blocks within a picture, based on the reference picture index that is used for that motion block. Because the reference picture index is already transmitted in the compressed video bitstream, the additional overhead to adapt the weighting factor on a motion block basis is dramatically reduced. All motion blocks that are coded with respect to the same reference picture apply the same weighting factor to the reference picture prediction.

In the Joint Model ("JM") software of the JVT committee, an a posteriori method using rate distortion optimization is used for selection of motion vectors, macroblock partitioning, prediction mode, and reference picture indices. In this method, a range of allowable values for each of these choices is tested and a cost is  
5 determined for each choice. The choice that leads to the minimum cost is selected.

Motion estimation techniques have been widely studied. For each motion block of a picture being coded, a motion vector is chosen that represents a displacement of the motion block from a reference picture. In an exhaustive search method within a search region, every displacement within a pre-determined range of  
10 offsets relative to the motion block position is tested. The test includes calculating the sum of the absolute difference ("SAD") or mean squared error ("MSE") of each pixel in the motion block in the current picture with the displaced motion block in a reference picture. The offset with the lowest SAD or MSE is selected as the motion vector. Numerous variations on this technique have been proposed, such as three-  
15 step search and rate-distortion optimized motion estimation, all of which include the step of computing the SAD or MSE of the current motion block with a displaced motion block in a reference picture.

Computational costs for determining motion vectors and adaptive reference picture weighting factors can be reduced by using an iterative process, while still  
20 selecting motion vectors and weighting factors that are able to achieve high compression efficiencies. An exemplary embodiment motion vector and weighting factor determination process is described assuming that a single weighting factor is applied to the entire reference picture, although the principles of the invention should not be construed as being so limited. The process could also be applied over smaller  
25 regions of the picture, such as slices, for example. In addition, although one exemplary embodiment is described as using only a single reference picture, the principles may also be applied to multiple reference picture prediction and to bi-predictive pictures.

Calculation of the motion vector for a motion block can typically best be done  
30 when the weighting factor to be used is known. In an exemplary embodiment, an estimate of the weighting factor is formed, using the reference picture and the current picture pixel values. The weighting factor may be limited to a number of bits of resolution. If the weighting factor is very close to 1, there is no need to consider the



weighting factor in the motion estimation process, and normal motion estimation can be done with the weighting factor assumed to be equal to 1. Otherwise, the weighting factor estimate is applied to the reference picture. Motion estimation is then performed using any method which calculates SAD or MSE, but with the SAD or MSE calculation performed between the current picture motion block and the displaced motion block in the weighted version of the reference picture, rather than the un-weighted reference picture. The estimation of the weighting factor can be refined after the motion vectors have been selected, if necessary.

The current motion vectors are applied to the weighted reference picture to form the weighted, motion compensated reference picture. A difference measure between the weighted, motion compensated reference picture and the current picture is computed. If the difference measure is lower than a threshold, or lower than the previous best difference measure, the process is complete, and the current candidate motion vectors and weighting factor are accepted.

If the difference measure is higher than some threshold, the weighting factor can be refined. In this case, a motion compensated but un-weighted reference picture is formed based on the current candidate motion vectors. The weighting factor estimate is refined using the motion compensated reference picture and the current picture, rather than using the un-compensated reference picture, as was done in forming the initial estimate of the weighting factor.

In one embodiment, the initial estimate of the weighting factor,  $w$ , is the ratio between the average value of the pixels in the current picture,  $cur$ , divided by the average value of the pixels in the reference picture,  $ref$ , where:

$$w = \text{avg}(cur) / \text{avg}(ref) \quad (4)$$

The refinement estimates are the ratio between the average of pixels in the current picture and the average of pixels in the motion compensated reference picture,  $mcref$ , where:

$$w = \text{avg}(eur) / \text{avg}(mcref) \quad (5)$$

The difference measure *diff* is the absolute value of the average of pixel differences between the current picture, *cur*, and the weighted motion compensated reference picture, *wmcref*, where:

$$\text{diff} = | \sum \text{cur} - \text{wmcref} | \quad (6)$$

In another embodiment, the difference measure is the sum of the absolute differences of the pixels in the current picture and in the weighed motion compensated reference picture, where:

$$\text{diff} = \sum | \text{cur} - \text{wmcref} | \quad (7)$$

When block-based motion estimation is performed, the same pixel in a reference picture is used for numerous SAD calculations. In an exemplary embodiment during the motion estimation process, once a weighting factor has been applied to a pixel in a reference picture, the weighted pixel is stored, in addition to the normal pixel. The storage may be done either for a region of the picture, or for the entire picture.

The weighted reference picture values may be clipped to be stored with the same number of bits as an unweighted reference, such as 8 bits, for example, or may be stored using more bits. If clipping is performed for the motion compensation process, which is more memory efficient, the weighting factor is reapplied to the reference picture for the actual selected motion vector, the difference is calculated using additional bits, and the clipping is performed after the difference in order to avoid mismatch with a decoder, which might otherwise occur if the decoder does not perform clipping after the weighting factor is applied.

When multiple reference pictures are used to encode a picture, a separate weighting factor can be calculated for each reference picture. During motion estimation, a motion vector and a reference picture index are selected for each motion block. For each iteration of the process, motion vectors and weighting factors are found for each reference picture.

In a preferred embodiment, during motion estimation, the best reference picture for a given motion block is determined. Calculation of the difference measure

is done separately for each reference picture, with only those motion blocks that use that reference picture being used in the calculation. Refinement of the weighting factor estimate for a given reference picture also uses only those motion blocks that are coded using that reference picture. For bi-predictive coding, weighting factors and motion vectors can be determined separately for each of the two predictions, which will be averaged together to form the averaged prediction.

The principles of the present invention can be applied to many different types of motion estimation algorithms. When used with hierarchical approaches, the iteration of weighting factor selection and motion vector selection can be used with any level of the motion estimation hierarchy. For example, the iterative approach could be used with integer picture element ("pel") motion estimation. After the weighting factor and integer motion vectors are found using the provided iterative algorithm, the sub-pel motion vectors may be found without requiring another iteration of the weighting factor selection.

These and other features and advantages of the present invention may be readily ascertained by one of ordinary skill in the pertinent art based on the teachings herein. It is to be understood that the principles of the present invention may be implemented in various forms of hardware, software, firmware, special purpose processors, or combinations thereof.

Most preferably, the principles of the present invention are implemented as a combination of hardware and software. Moreover, the software is preferably implemented as an application program tangibly embodied on a program storage unit. The application program may be uploaded to, and executed by, a machine comprising any suitable architecture. Preferably, the machine is implemented on a computer platform having hardware such as one or more central processing units ("CPU"), a random access memory ("RAM"), and input/output ("I/O") interfaces. The computer platform may also include an operating system and microinstruction code. The various processes and functions described herein may be either part of the microinstruction code or part of the application program, or any combination thereof, which may be executed by a CPU. In addition, various other peripheral units may be connected to the computer platform such as an additional data storage unit and a printing unit.

It is to be further understood that, because some of the constituent system components and methods depicted in the accompanying drawings are preferably implemented in software, the actual connections between the system components or the process function blocks may differ depending upon the manner in which the present invention is programmed. Given the teachings herein, one of ordinary skill in the pertinent art will be able to contemplate these and similar implementations or configurations of the present invention.

Although the illustrative embodiments have been described herein with reference to the accompanying drawings, it is to be understood that the present invention is not limited to those precise embodiments, and that various changes and modifications may be effected therein by one of ordinary skill in the pertinent art without departing from the scope or spirit of the present invention. All such changes and modifications are intended to be included within the scope of the present invention as set forth in the appended claims.

CLAIMS

1. A video encoder (200, 300) for encoding video signal data for at least one cross-fade picture disposed temporally between a fade-out start picture and a fade-in end picture, which are used as reference pictures for coding the at least one cross-fade picture, the encoder comprising:
  - a reference picture weighting applicator (292, 392); and
  - a reference picture weighting factor unit (272, 372) in signal communication with the reference picture weighting applicator for assigning weighting factors corresponding to each of the fade-out start picture and the fade-in end picture, respectively, for coding the at least one cross-fade picture.
2. A video encoder as defined in Claim 1, further comprising a motion compensation unit (290, 390) in signal communication with the reference picture weighting applicator for providing at least one of a motion compensated fade-out start picture and a motion compensated fade-in end picture responsive to the reference picture weighting factor unit for coding the at least one cross-fade picture.
3. A video encoder as defined in Claim 2, further comprising a reference picture store (270, 370) in signal communication with each of the reference picture weighting factor unit and the motion compensation unit for storing each of the fade-out start picture and the fade-in end picture.
4. A video encoder as defined in Claim 2 wherein the reference picture weighting applicator applies a weighting factor selected by the reference picture weighting factor unit to at least one of the motion compensated fade-out start picture and the motion compensated fade-in end picture.
5. A video encoder as defined in Claim 4 usable with bi-predictive picture predictors, the encoder further comprising prediction means for forming first and second predictors from the weighted and motion compensated fade-out start and fade-in end pictures, respectively.

6. A video encoder as defined in Claim 5 wherein the weighted and motion compensated fade-out start and fade-in end pictures, respectively, are each from opposite directions relative to all of the at least one cross-fade pictures.

5           7. A video encoder as defined in Claim 1, further comprising a motion estimation unit (380) in signal communication with the reference picture weighting factor unit for providing motion estimation responsive to weighting factor in an explicit mode of operation.

10           8. A video encoder as defined in Claim 2, further comprising a summing unit (394) in signal communication with the reference picture weighting factor unit for applying an offset to the weighted motion compensated reference picture in an explicit mode of operation.

15           9. A method (700) for encoding cross-fades between pictures, the method comprising:

identifying pictures for which a cross-fade is defined;

determining (714,716) appropriate end-points from pictures for which said cross-fade is defined; and

20           encoding (718,720) said end-points prior to encoding (722) at least one picture intermediate to said end-points.

10. A method as defined in Claim 9 wherein said end-points from pictures for which said cross-fade is defined are used as reference pictures when encoding at  
25           least one picture intermediate to said end-points.

11. A method as defined in Claim 9, further comprising:

receiving a substantially uncompressed fade-out start picture; receiving a substantially uncompressed fade-in end picture;

30           assigning a weighting factor for the at least one - picture corresponding to the fade-out start picture; and

assigning a weighting factor for the at least one - picture corresponding to the fade-in end picture.

12. A method as defined in Claim 11, further comprising:

computing motion vectors corresponding to the difference between the at least one cross-fade picture and at least one of the fade-out start picture and the fade-in end picture;

5        motion compensating the at least one of the fade-out start picture and the fade-in end picture in correspondence with the motion vectors;

      multiplying the motion compensated at least one of the fade-out start picture and the fade-in end picture by the assigned weighting factor, respectively, to form at least one weighted motion compensated reference picture; and

10       subtracting the at least one weighted motion compensated reference picture from the at least one cross-fade picture; and encoding a signal indicative of the difference between the at least one cross-fade picture and the at least one weighted motion compensated reference picture.

15       13. A method as defined in Claim 12 wherein exactly two reference pictures are used, the exactly two reference pictures comprising the pre-coded fade-out start picture, FP0, and the fade-in end picture, FP1.

14. A method as defined in Claim 13, further comprising:

20       combining the motion compensated fade-out start picture with the motion compensated fade-in end picture prior to subtracting from the at least one cross-fade picture.

15. A method as defined in Claim 12 wherein computing motion vectors  
25       comprises:

      testing within a search region for every displacement within a pre-determined range of offsets relative to the at least one cross-fade picture;

      calculating at least one of the sum of the absolute difference and the mean squared error of each pixel in the at least one cross-fade picture with a motion  
30       compensated reference picture; and

      selecting the offset with the lowest sum of the absolute difference and mean squared error as the motion vector.

16. A method as defined in Claim 12 wherein computing motion vectors comprises:

testing within a search region for every displacement within a pre-determined range of offsets relative to the at least one cross-fade picture;

5 calculating at least one of the sum of the absolute difference and the mean squared error of each pixel in the at least one cross-fade picture with a first motion compensated reference picture corresponding to the fade-out start picture;

selecting an offset with the lowest sum of the absolute difference and mean squared error as the motion vector for the fade-out start picture;

10 calculating at least one of the sum of the absolute difference and the mean squared error of each pixel in the image block with a second motion compensated reference picture corresponding to the fade-in end picture; and

selecting an offset with the lowest sum of the absolute difference and mean squared error as the motion vector for the fade-in end picture.

15

17. A method as defined in Claim 11 wherein the weighting factors for the fade-out start picture and the fade-in end picture, respectively, are each responsive to the relative distance between the at least one cross-fade picture and the fade-out start picture or the fade-in end picture, respectively, in an implicit mode of operation.

20

18. A video CODEC comprising an encoder as defined in Claim 1 and a decoder (500) for decoding video signal data for a cross-fade picture relative to each of a fade-out start picture and a fade-in end picture to predict the cross-fade picture, the decoder comprising a reference picture weighting factor unit (580) having an  
25 output for determining weighting factors corresponding to each of the fade-out start picture and the fade-in end picture.

19. A video CODEC as defined in Claim 18 wherein the reference picture weighting factor unit has a second output for determining offsets corresponding to  
30 each of the fade-out start picture and the fade-in end picture.



20. A video CODEC as defined in Claim 18, further comprising a variable length decoder (510) in signal communication with the reference picture weighting factor unit for providing indices corresponding to each of the fade-out start picture and the fade-in end picture to the reference picture weighting factor unit.

5

21. A video CODEC as defined in Claim 18, further comprising a motion compensator (560) in signal communication with the reference picture weighting factor unit for providing motion compensated reference pictures responsive to the reference picture weighting factor unit.

10

22. A video CODEC as defined in Claim 21, further comprising a reference picture weighting applicator (570) in signal communication with the motion compensator and the reference picture weighting factor unit for applying a weighting factor to each motion compensated reference picture.

15

23. A video CODEC as defined in Claim 21, further comprising an adder (590) in signal communication with the motion compensator and the reference picture weighting factor unit for applying an offset to each motion compensated reference picture.

20

24. A video CODEC as defined in Claim 18 wherein the video signal data is streaming video signal data comprising block transform coefficients.

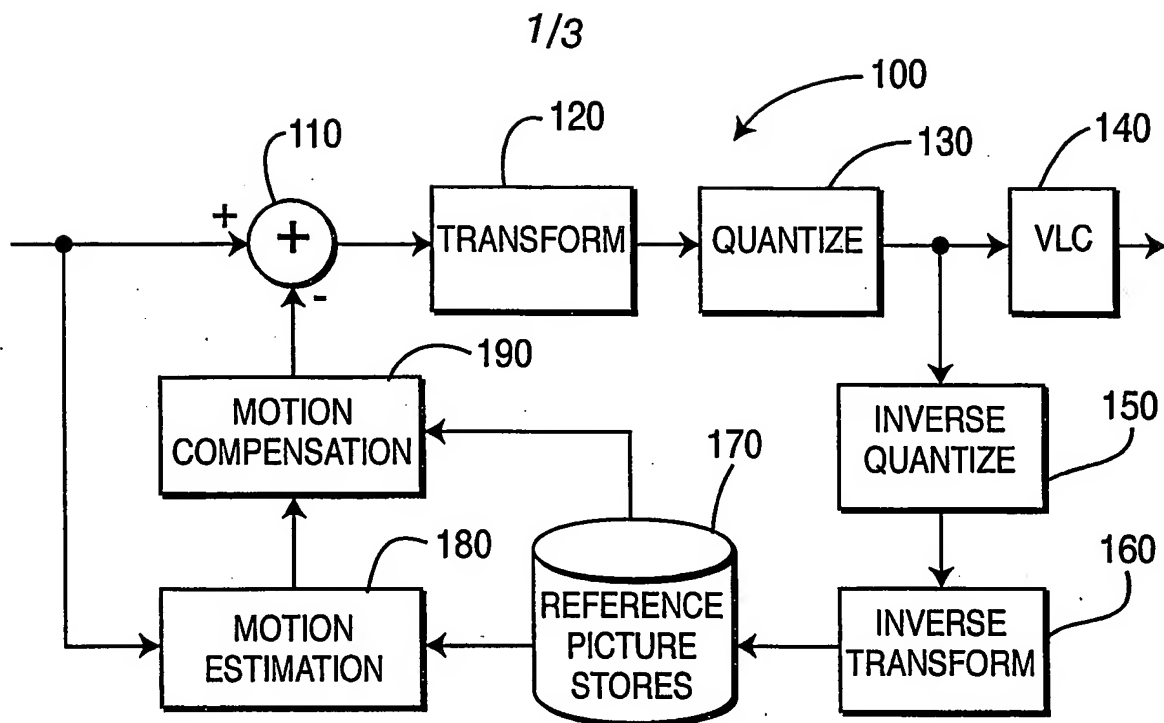
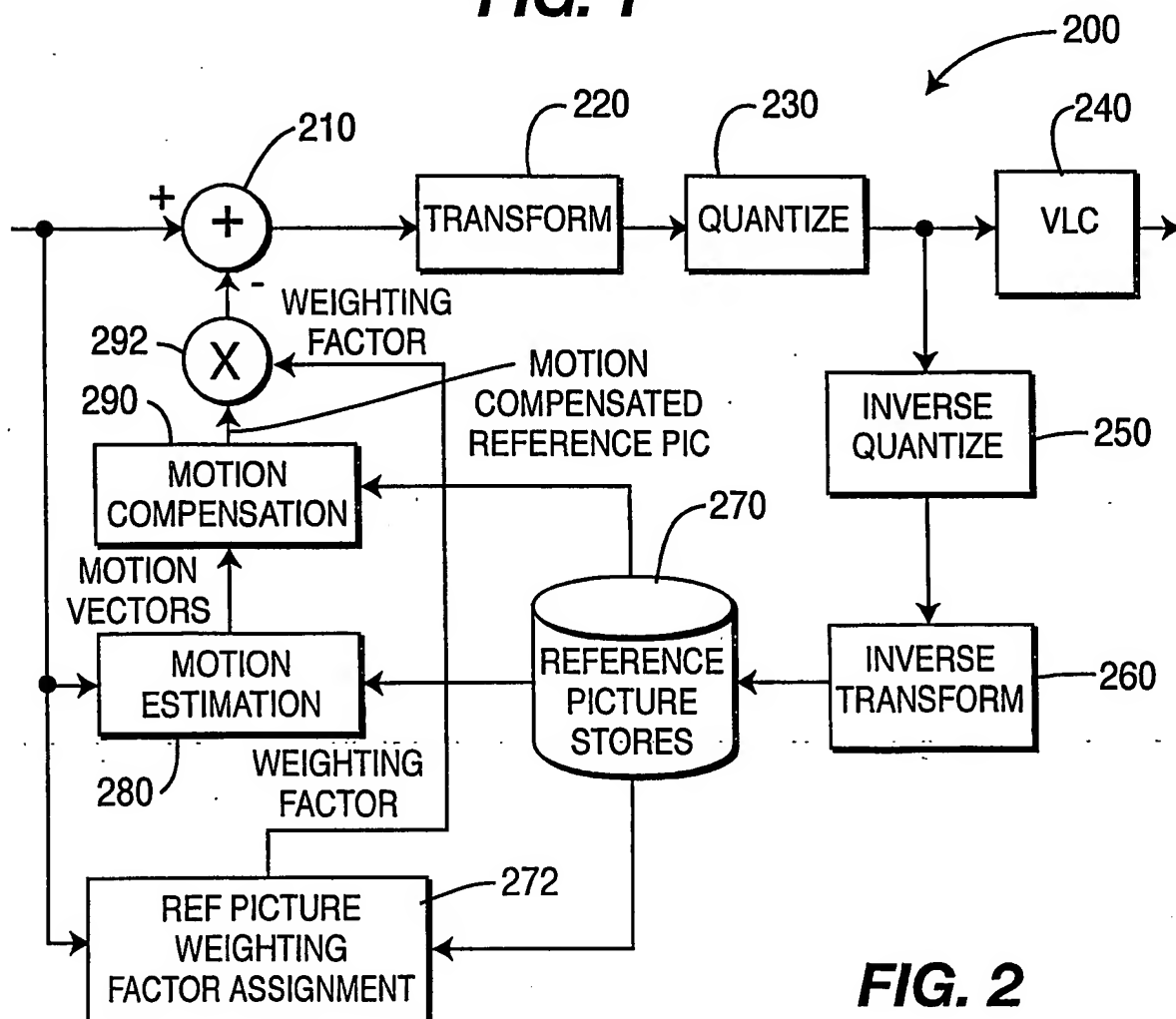
25. A video CODEC as defined in Claim 18 usable with bi-predictive picture predictors, the decoder further comprising:

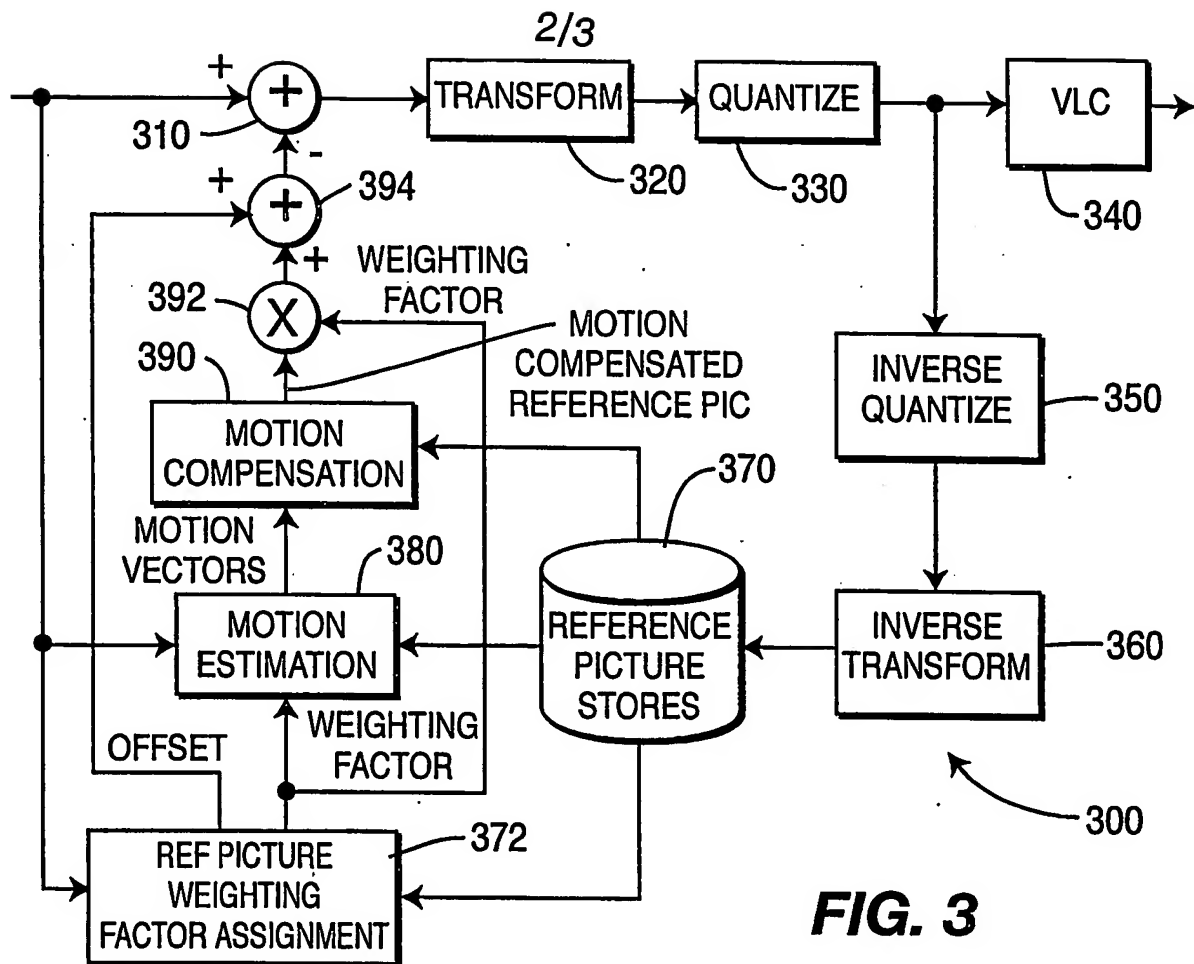
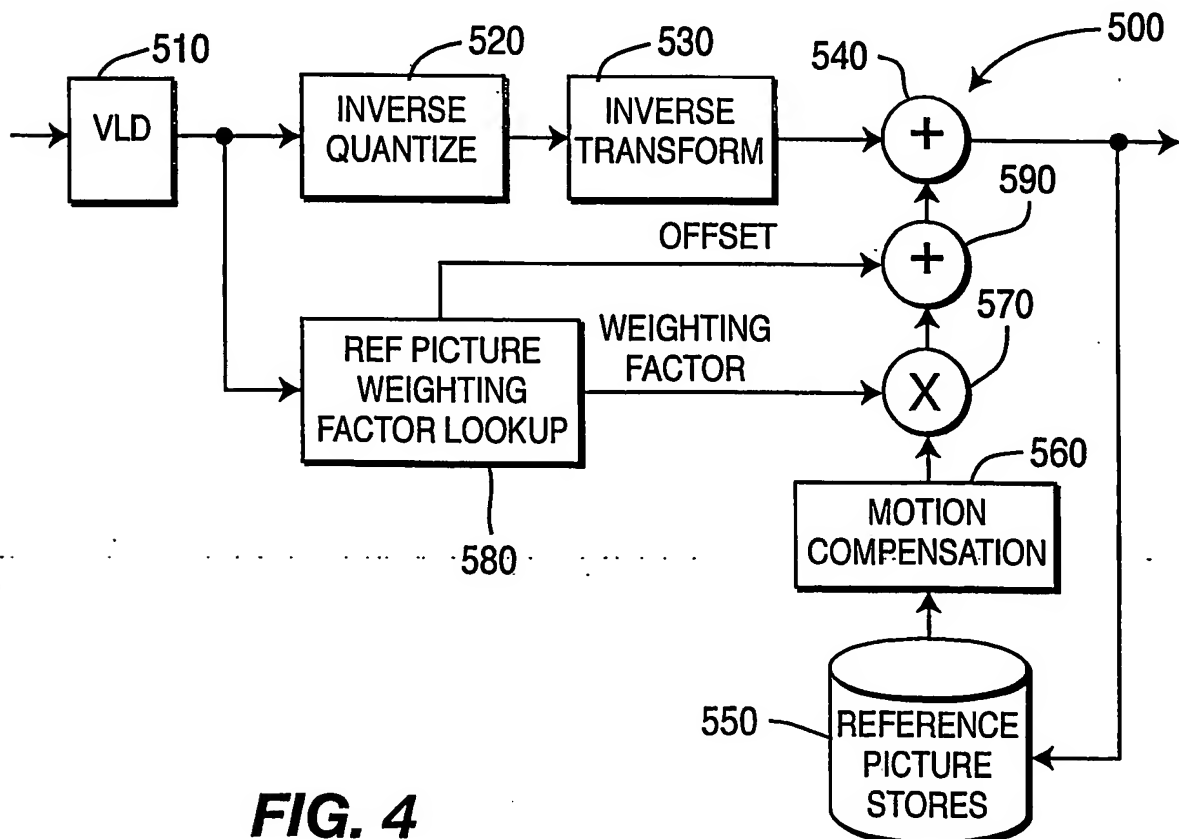
25

prediction means for forming first and second predictors from two different reference pictures;

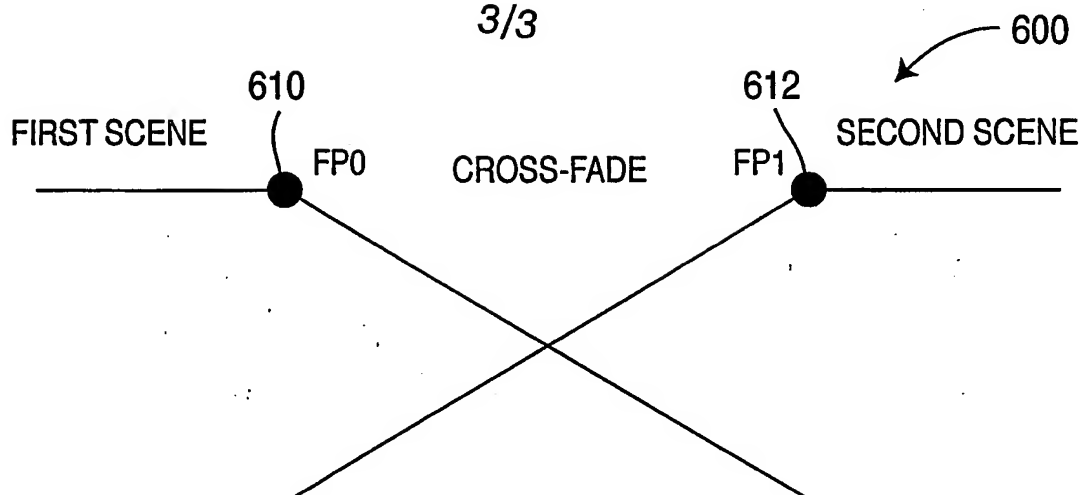
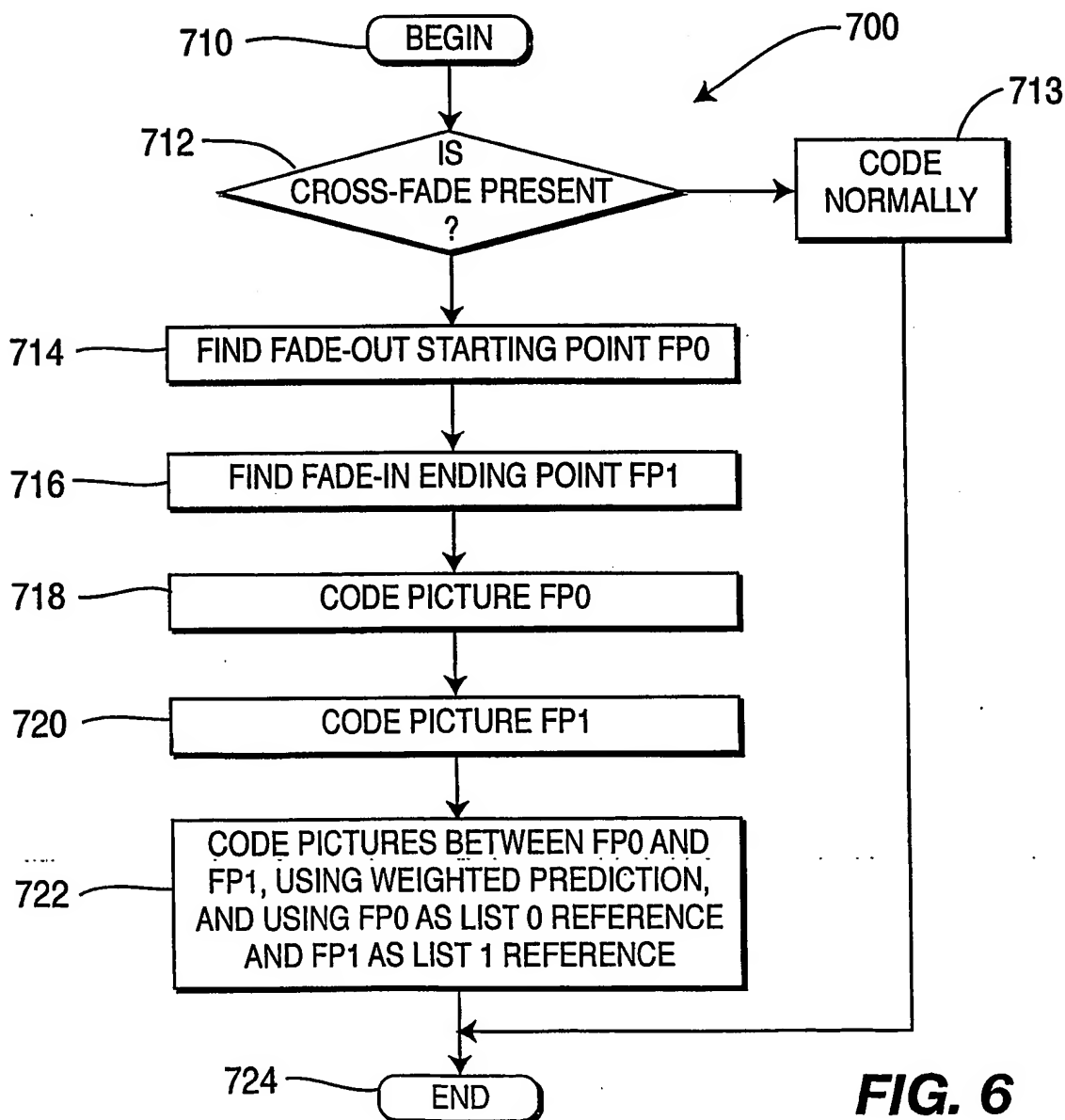
averaging means for averaging the first and second predictors together using their corresponding weighting factors to form a single averaged predictor.

30

**FIG. 1****FIG. 2**

**FIG. 3****FIG. 4**

3/3

**FIG. 5****FIG. 6**